Protein Science (1994), *3*:1344–1346. Cambridge University Press. Printed in the USA.

FOR THE RECORD

A new family of carbon–nitrogen hydrolases

PEER BORK[1] AND EUGENE V. KOONIN[2]

[1] European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, and Max-Delbrück-Center, 13125 Berlin-Buch, Germany
[2] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894

(RECEIVED April 18, 1994; ACCEPTED May 24, 1994)

Abstract: Using computer methods for database search and multiple alignment, statistically significant sequence similarities were identified between several nitrilases with distinct substrate specificity, cyanide hydratases, aliphatic amidases, β-alanine synthase, and a few other proteins with unknown molecular function. All these proteins appear to be involved in the reduction of organic nitrogen compounds and ammonia production. Sequence conservation over the entire length, as well as the similarity in the reactions catalyzed by the known enzymes in this family, points to a common catalytic mechanism. The new family of enzymes is characterized by several conserved motifs, one of which contains an invariant cysteine that is part of the catalytic site in nitrilases. Another highly conserved motif includes an invariant glutamic acid that might also be involved in catalysis.

Keywords: enzyme evolution; homology search; nitrogen metabolism

Utilization of nitrogen compounds usually involves several reduction steps. The final pathway frequently is the NH_4^+ assimilation or transfer of NH_4^+ to various intermediates such as keto acids (for review, see Willison, 1993, and references therein). We report here significant sequence conservation among numerous enzymes that are involved in nitrogen metabolism and that cleave nitriles as well as amides (Fig. 1; Table 1). Based on their sequence similarity and on the reactions they catalyze, these enzymes can be classified into 6 functionally distinct groups.

1. Nitrilases cleave various nitriles into the corresponding acids and ammonia. The family described here includes enzymes with specific substrates such as bromoxynitrile, arylacetone, and indole-3-acetonitrile, but also wide-spectrum nitrilases (e.g., acting on aromatic or aliphatic nitriles). Prokaryotic nitrilases are mainly involved in utilization of various nitrogen sources (Nagasawa et al., 1990; Kobayashi et al., 1992); plant nitrilases participate in the production of the hormone indole-3-acetic acid (Kobayashi et al., 1993).
2. Cyanide hydratase of pathogenic fungi detoxifies HCN that is released by their hosts, cyanogenic plants, after injury. Its sequence is very similar to those of nitrilases (Wang & Van Etten, 1992) and is actually more closely related to prokaryotic nitrilases than the latter are to eukaryotic ones (Fig. 1).
3. Aliphatic amidases enable prokaryotes to use acetamides as both carbon and nitrogen source (see Soubrier et al., 1992, and references therein).
4. β-Alanine synthase (*N*-carbamoyl-β-alanine amido hydrolase) catalyzes the last step of pyrimidine catabolism; the resulting β-alanine can be further converted into pyruvate but can also be incorporated into various pathways. In addition to the characterized enzyme from rat (Kvalnes-Krick & Traut, 1993), our database search revealed a closely related (50% sequence identity over 98 residues) expressed sequence tag (EST) from *Caenorhabditis elegans* (EMBL accession no. Z14933).
5. AdgA (for ammonia-dependent growth) from *Rhodobacter* species appears to be essential for using various amino acids as nitrogen sources; mutations in this gene lead to the requirement for ammonia in the medium for growth (Willison, 1993). AdgA is the only protein of this sequence family with an additional domain (Willison, 1993), a putative ATP pyrophosphatase (data not shown).
6. An open reading frame (ORF5) in the *agr*-operon of *Staphylococcus* species (Vandenesch et al., 1993; J. Kornblum, S.J. Projan, B.N. Kreiswirth, S.L. Mogazeh, W. Eisner, H. Ross, R.P. Novick, unpubl.: EMBL accession no. X52543; Agr5 in Fig. 1) comprise yet another distinct group and probably possess an as yet uncharacterized enzymatic activity.

The sequence similarity within the 6 groups is very high, whereas it ranges only between 12 and 24% amino acid identity between the groups (Fig. 1). Database searches performed using the Blastp program (Altschul et al., 1990) indicated a very low probability (*P*-value) of matching by chance for nitrilases versus cyanide hydratases (below 10^{-20}) and for ORF5 in the *agr*-operon versus aliphatic amidases (below 10^{-8}). The *P*-value of

Reprint requests to: Peer Bork, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany; e-mail: bork@embl-heidelberg.de.

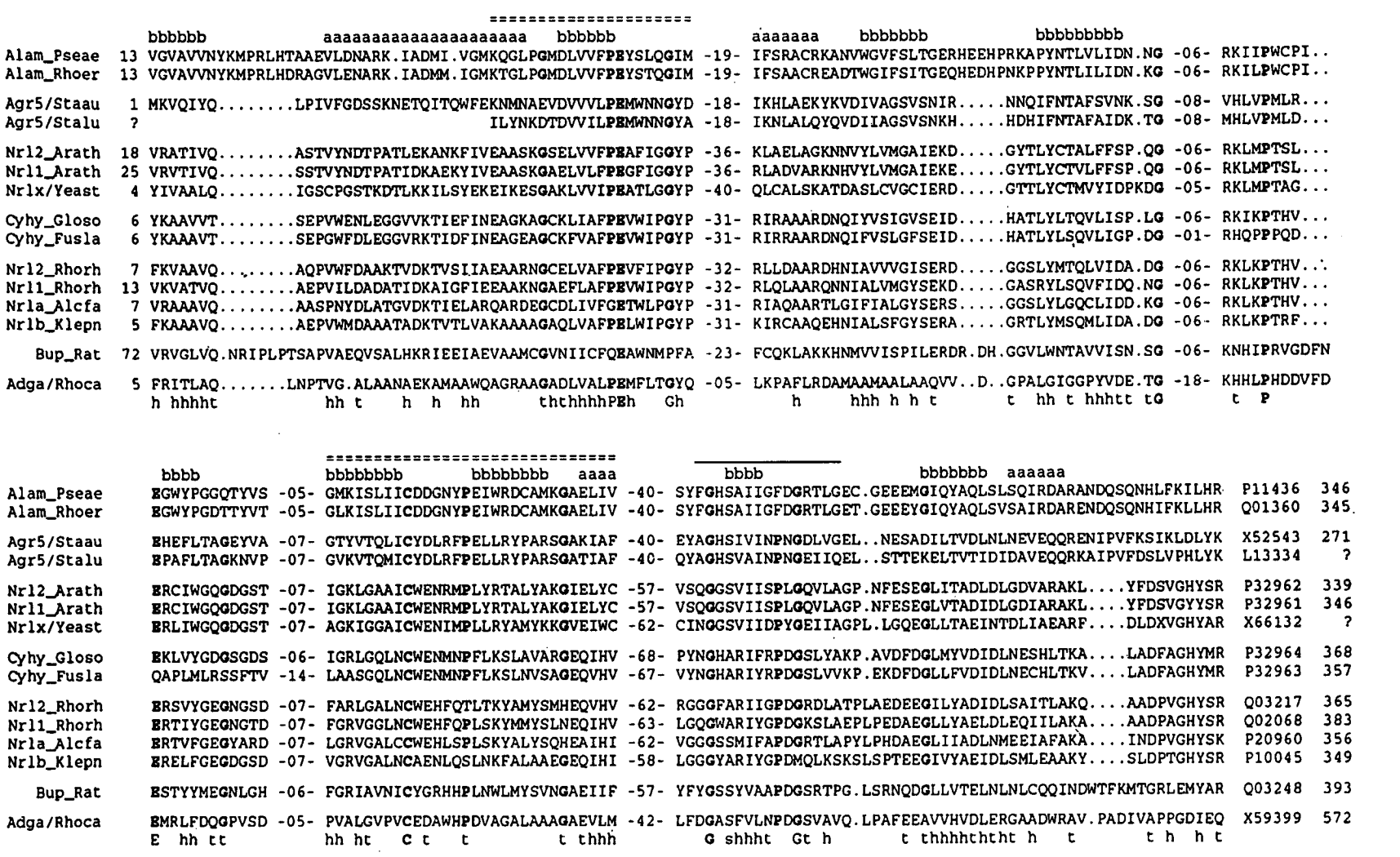

Fig. 1. Multiple alignment of the family of C-N bond-cleaving enzymes. SWISS-PROT codes (Bairoch & Boeckmann, 1993) were used to identify proteins if available (underscores within the abbreviations): Alam, aliphatic amidase; Arg5, ORF5 in the *arg* operon; Nrl, nitrilase; Cyhy, cyanide hydratase; Bup, β-alanine synthase; Adga, adgA gene product. The left column shows the positions within the respective sequences. Top line, secondary-structure elements as predicted using the PhD program (Rost & Sander, 1993): a, α-helix; b, β-strand. Bottom line, consensus of the alignment: capitals, amino acids conserved in at least all but 3 sequences; h, hydrophobic positions; t, turnlike or polar positions. Numbers within the alignment indicate the number of amino acids in between the displayed blocks. The total length of the sequences is given in the rightmost column preceded by the database accession numbers. The 3 motifs with statistically significant conservation as revealed by the MACAW program are overlined; the 2 that have been used for PROPAT searches are double overlined. In order to assemble the Nrlx/Yeast sequence (from EMBL accession no. X66132), a base had to be replaced (stop codon > tryptophan) and another one deleted (frameshift). Modifications were indicated by the BLASTX program (Gish & States, 1993) that translates a DNA region into all 6 possible reading frames and compares it with protein sequence databases. Members of the family were identified by iterative database searches using programs of the BLAST series followed by pattern searches (for details see, e.g., Koonin et al., 1994).

Table 1. *Summary of the carbon-nitrogen hydrolase family*

Enzyme	EC number	Identical subunits	Chemical reaction	References
β-Alanine synthase	EC 3.5.1.6	Hexamer	*N*-carbamoyl-β-alanine + H_2O = β-alanine + CO_2 + NH_3	Kvalnes-Krick & Traut, 1993
Cyanide hydratase	EC 4.2.1.66	Multimer	Hydrogen cyanide + H_2O = formamide (CHO-NH_2)	Cluness et al., 1993
Aliphatic amidase	EC 3.5.1.4	Tetramer	Monocarboxylic acid amide + H_2O = monocarboxylate + NH_3	Soubrier et al., 1992
Nitrilases	EC 3.5.5.1	Hexamer, dimer	Nitrile + H_2O = carboxylate + NH_3	Kobayashi et al., 1992
AdgA	?	?	X-NH_2 + ATP + H_2O = Y + AMP + PP + NH_3	Willison, 1993

7.7×10^{-4} was observed for the alignment between β-alanine synthase and one of the fungal cyanide hydratases. In contrast, sequences from other subsets of the family (Fig. 1) showed only very low, not statistically significant, similarity to each other in these initial searches. Nevertheless, the fact that the segments detected by Blastp included the same regions that are most highly conserved in each of the subsets prompted further investigation of these marginal similarities.

Multiple alignment with the ClustalW program (Higgins et al., 1992; D. Higgins, J. Thompson, & T. Gibson, unpubl.) revealed an overall similarity of all proteins (Fig. 1). The alignment was confirmed with the MACAW program (Schuler et al., 1991), which calculated for 3 blocks (overlined in Fig. 1) a probability of matching by chance below 10^{-7} (Fig. 1). For the 2 most conserved motifs (double overlined in Fig. 1), property pattern database searches were carried out (PROPAT program; Rohde & Bork, 1993). We found that the patterns were unique for this family, i.e., no additional sequences with comparable level of similarity to the conserved patterns were detected.

The conserved blocks span almost the entire length of the proteins in the new family (with the only exception of AdgA that contains an additional domain), which is highly suggestive of a common topology. This is supported by secondary structure predictions (Rost & Sander, 1993) that give similar results for different groups (Fig. 1).

Although one has to be very cautious in transferring functional information between distantly related sequences (Bork et al., 1994), a few generalizations can be made. One of the conserved motifs shared by all the proteins of the family contains an invariant cysteine (Fig. 1). This residue has been shown to be involved in the active site of very different nitrilases (Kobayashi et al., 1992, 1993). It appears likely that this cysteine plays the same role in all members of this family of enzymes. The most conserved motif (the probability of chance occurrence is 7.9×10^{-18} as calculated using the MACAW algorithm) is located toward the N-terminus of the proteins. It contains an invariant glutamic acid that follows a very hydrophobic, probably interior β-strand (Fig. 1). The glutamic acid residue might thus be another direct participant in the cleavage reaction.

An interesting feature of the family described here is the strong hydrophobicity within the conserved regions. The hydrophobicity might be due to subunit contacts and indeed all of the characterized proteins appear to form multimers (Table 1). An alternative explanation is a tightly packed β-sheet with the hydrophobic β-strands located in the core and shielded from the protein surface.

The family described here consists of functionally diverse proteins such as nitrilases, amidases, cyanide hydratase, and β-alanine synthase. On the other hand, sequences are available for several other nitrilases and amidases that appear to be completely unrelated to the members of the family delineated here. Thus, enzymes involved in the metabolism of various organic nitrogen compounds seem to be yet another example of convergent evolution of proteins toward similar enzymatic activity.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman D. 1990. Basic local alignment search tool. *J Mol Biol 215*:403–410.

Bairoch A, Boeckmann B. 1993. The SWISS-PROT protein sequence data bank, recent developments. *Nucleic Acids Res 21*:3093–3096.

Bork P, Ouzounis C, Sander C. 1994. From genome sequences to protein function. *Curr Opin Struct Biol 4*:393–403.

Cluness MJ, Turner PD, Clements E, Brown DT, O'Reilly C. 1993. Purification and properties of cyanide hydratase from *Fusarium lateritium* and the analysis of the *chy1* gene. *J Gen Microbiol 139*:1807–1815.

Gish W, States DJ. 1993. Identification of protein coding regions by database similarity search. *Nature Genet 3*:266–272.

Higgins D, Bleasby AJ, Fuchs R. 1992. ClustalV: Improved software for multiple alignment. *CABIOS 8*:189–191.

Kobayashi M, Izui H, Nagasawa T, Yamada H. 1993. Nitrilase in biosynthesis of the plant hormone indole-3-acetic acid from indole-3-acetonitrile. *Proc Natl Acad Sci USA 90*:247–251.

Kobayashi M, Yanaka N, Nagasawa T, Yamada H. 1992. Primary structure of an aliphatic nitrile-degrading enzyme, aliphatic nitrilase from *Rhodococcus rhodochrous* K22 and expression of its gene and identification of its active site residue. *Biochemistry 31*:9000–9007.

Koonin EV, Bork P, Sander C. 1994. Yeast chromosome III: New gene functions. *EMBO J 13*:493–503.

Kvalnes-Krick KL, Traut TW. 1993. Cloning, sequencing and expression of a cDNA encoding β-alanine synthase from rat liver. *J Biol Chem 268*: 5686–5693.

Nagasawa T, Mauger J, Yamata H. 1990. A novel nitrilase, arylacetonitrilase of *A. faecalis* JM3. Purification and characterization. *Eur J Biochem 194*:765–772.

Rohde K, Bork P. 1993. A fast, sensitive pattern-matching approach for protein sequences. *CABIOS 9*:183–189.

Rost B, Sander C. 1993. Prediction of protein secondary structure at better than 70%. *J Mol Biol 232*:584–599.

Schuler GD, Altschul SF, Lipman D. 1991. A workbench for multiple sequence alignment construction and analysis. *Proteins Struct Funct Genet 9*:180–190.

Soubrier F, Levy-Schil S, Mayaux JF, Petre D, Arnaud A, Crouzet J. 1992. Cloning and primary structure of the wide-spectrum amidase from *Brevibacterium* sp. R312: High homology to the *amiE* product from *P. aeruginosa*. *Gene 116*:99–104.

Vandenesch F, Projan SJ, Kreiswirth B, Etienne J, Novick RP. 1993. *Agr*-related sequences in *Staphylococcus lugdunensis*. *FEMS Microbiol Lett 111*:115–122.

Wang P, Van Etten HD. 1992. Cloning and properties of a cyanide hydratase gene from the phytopathogenic fungus *Gloeocercospora sorghi*. *Biochem Biophys Res Commun 187*:1048–1054.

Willison JC. 1993. Biochemical genetics revisited: The use of mutants to study carbon and nitrogen metabolism in the photosynthetic bacteria. *FEMS Microbiol Rev 104*:1–38.